Data Dictionary: A Tool to Expedite Information Systems Integration

Anne Jimenez, Karen Dicksen, Wanda Noble, Susan Butler, Kurt Gunther,
Joyce Niland, Ph.D.
Departments of Biostatistics and Information Services
City of Hope National Medical Center
Duarte, California

The City of Hope (COH) National Medical Center is a specialty institution which conducts clinical and basic science research into the etiology, prevention and treatment of catastrophic disease, principally cancer. COH was awarded an Integrated Advanced Information Management Systems (IAIMS) planning grant to develop a comprehensive strategic plan for an institution-wide integrated information system (CRIIS: the Clinical and Research Integrated Information System). One of the primary issues which CRIIS seeks to address is the integration of the in-house-developed and vendor-supplied applications and maximize the utility of information systems to research scientists, clinicians and administrators. A master data dictionary is envisioned as a key tool which will facilitate and expedite the system integration process.

Initially, the master data dictionary will be used to illuminate the relationship between existing systems: identify what data is stored; where it is stored; how data flows between systems; and who enters, updates and uses the data. Redundant data—a potential source of inconsistency—will be identified. Developers of new systems will be able to more easily identify existing sources of data, reducing manual input and duplication of data and the potential for error. Beyond clarifying existing data relationships, the master data dictionary will expedite future development by making it easier for developers to identify sources of data, to locate expertise with targeted development tools, and to assess the appropriateness of development tools.

One of the key requirements for the successful implementation of a master data dictionary is accessibility. COH is committed to a campus-wide fiber optic network and already has in place more than 30,800 feet of fiber consisting of 144 strands of fiber per cable, and linking all 63 buildings. Also in place are committees and task forces under the CRIIS project to manage information resources and ensure that decisions regarding information systems are consistent with the goal of integrating systems. Selection of software for the data dictionary will reflect the need to make the data dictionary accessible to users from diverse hardware and software platforms.

The role of the master data dictionary will not be limited to the documentation of system relationships: it will showcase the spirit and intent of the CRIIS project. For example, an on-line interface will be developed first that will be used to gather data about existing systems. Interviews with system developers and users will be done on-line, whenever possible from the interviewee's office. This will demonstrate the commitment to accessibility and it will allow users to see the data dictionary system in action. Whenever possible, other under-construction features of the data dictionary will be demonstrated during the interview.

The implementation strategy will be a sandwich type construction. The top and bottom layers of the system will be developed in the early stages: the top layer will be the on-line data entry system which will be used during the interview of system users and developers to gather information about existing systems; the bottom layer will be the design and construction of the data dictionary database. The middle layers will be filled in as required to load the data dictionary and to access the information. Prototype data from systems will be entered as early as possible to provide a base for demonstrations. This will allow users to also give feedback and suggestions during the interviews.

Ultimately, it is our goal to implement the data dictionary in a manner that is supportive of the CRIIS project and maximizes not only the benefits to its potential users, but also maximizes the number of potential users, and expedites the CRIIS project itself.

0195—4210/96/$5.00